

# Filosofare artificiale: minaccia o opportunità per il pensiero?

di: [ANDREA COLAMEDICI](#) e [MAURA GANCITANO](#)

## Sommario

L'avvento dell'intelligenza artificiale (AI) sta ridefinendo numerosi aspetti della società umana, incluso il modo in cui pensiamo e ragioniamo intorno a questioni esistenziali complesse. Questo paper esplora l'impatto potenziale dell'AI sul processo filosofico e sul pensiero, analizzando le potenziali minacce e opportunità presentate dai sistemi di AI avanzati, in particolare dall'Intelligenza Artificiale Generativa, nel campo della filosofia. Attraverso un'analisi critica della letteratura esistente e considerazioni nel campo dell'etica, il paper intende esaminare come l'AI possa influenzare l'indagine filosofica, la produzione di idee e la natura stessa del ragionamento astratto. Possiamo pensare meglio grazie a tali strumenti? O siamo destinati a perdere gradualmente le nostre capacità cognitive, finendo col delegare l'intero processo razionale a un agente esterno, come Platone a suo tempo temeva accadesse con l'avvento della scrittura? A tale scopo saranno prese in considerazione le implicazioni etiche, epistemologiche e ontologiche di questa tecnologia emergente, attraverso una serie di "esperimenti filosofici" con AI generative quali ChatGPT 4o e Gemini, valutando il loro potenziale sia come strumenti per amplificare il pensiero umano sia come possibili fonti di omogeneizzazione e manipolazione intellettuale. Mentre l'AI presenta opportunità non di poco conto per espandere gli orizzonti della ricerca filosofica, offre anche sfide enormi per il pensiero e per il nostro stesso dici "umani". Affinché ciò accada è necessario un approccio equilibrato e consapevole, che sappia portare a massimizzare i benefici minimizzando i rischi.

## Abstract

The advent of artificial intelligence (AI) is redefining numerous aspects of human society, including the way we think and reason about complex existential questions. This paper explores the impact of AI on the philosophical process and free thought, analyzing the potential threats and opportunities presented by advanced AI systems, particularly Generative Artificial Intelligence, in the field of philosophy. Through a critical analysis of existing literature and considerations in the field of ethics, the paper aims to examine how AI can influence philosophical inquiry, the production of ideas, and the very nature of abstract reasoning. Can we think better thanks to such tools? Or are we destined to gradually lose our cognitive abilities, ending up delegating the entire rational process to an external agent, as Plato once feared would happen with the advent of writing? To this end, the ethical, epistemological, and ontological implications of this emerging technology will be considered through a series of "philosophical experiments" with generative AI such as ChatGPT 4o and Gemini, evaluating their potential both as tools to amplify human thought and as possible sources of intellectual homogenization and manipulation. While AI presents significant opportunities to expand the horizons of philosophical research, it also offers enormous challenges for thought and for our very notion of being "human". For this to happen, a balanced and conscious approach is needed, one that can maximize benefits while minimizing risks.

## Keywords

Artificial Intelligence (AI), Philosophy, Generative AI, Ethical implications, Cognitive abilities

## 1. Introduzione

L'intelligenza artificiale, una volta dominio esclusivo del genere fantascientifico e dei reami dell'immaginazione, è ora una realtà che permea un alveo crescente di aspetti della nostra vita quotidiana. Dai sistemi di raccomandazione che influenzano gran parte delle nostre scelte di consumo – quelli di Netflix, Amazon, Facebook, Spotify, per citarne alcuni – agli assistenti virtuali come Siri e Alexa che rispondono alle nostre domande, l'AI è rapidamente diventata un elemento onnipresente della società moderna, ramificando il proprio dominio a partire dai *device* e dalle nostre abitazioni intelligenti. Tuttavia, mentre

l'impatto in campi come l'economia, la medicina e l'intrattenimento è ampiamente riconosciuto e studiato, le sue implicazioni nelle discipline umanistiche sono ancora oggetto di dibattito e speculazione, e non c'è consenso su rischi e opportunità di tale utilizzo nell'ambito sopracitato.

La filosofia si trova ora, difatti, di fronte a una sfida inedita: come relazionarsi con sistemi non umani capaci di elaborare concetti complessi, generare argomentazioni e controargomentazioni logiche e persino produrre testi che imitano – talvolta in maniera molto efficace – il ragionamento filosofico? Questa situazione solleva la domanda fondamentale della nostra ricerca: l'AI rappresenta una minaccia o un'opportunità per il pensiero libero e l'indagine filosofica? La risposta è presto detta: entrambe le cose. È sia un'opportunità che una minaccia; ma in che misura? E quanto è utile correre il rischio della minaccia per l'ipotesi dell'opportunità? Il fatto stesso di arrivare a porsi domande del genere è un'occasione eccezionale per la filosofia stessa, che trova un eccellente campo d'azione in cui utilizzare il proprio corpus teoretico. Lungi dall'essere solo un oggetto di studio, quindi, l'AI diventa in questo modo uno specchio attraverso il quale la filosofia può esaminare se stessa, i suoi metodi e i suoi presupposti. Questo nuovo campo d'azione offre un'opportunità straordinaria per applicare (e potenzialmente espandere) il proprio sapere teoretico, mettendo alla prova i concetti consolidati nei millenni e sviluppandone di nuovi per affrontare le sfide di questo tempo. Inoltre, da tale incontro – o scontro – potrebbe emergere un'aspra ma efficace e circostanziata critica dell'AI da parte della filosofia. Questo processo critico, lungi dall'essere meramente distruttivo, potrebbe portare a una rivivificazione della disciplina filosofica stessa, anche qualora l'AI fosse considerata in toto un "nemico del pensiero". Confrontandosi con le sfide poste dall'AI, la filosofia potrebbe essere spinta a raffinare i propri strumenti concettuali, a riesaminare le proprie assunzioni fondamentali e a sviluppare nuove metodologie d'indagine, nonché a ritrovare uno spazio d'azione centrale nella società umana. In questo senso, l'AI non solo fornisce un nuovo oggetto di studio per la filosofia, ma agisce anche come un potenziale catalizzatore per l'innovazione e il rinnovamento della disciplina.

## **2. Fornire domande più che risposte**

### **2.1. Le questioni chiave**

Quella in esame è un'interazione di certo non priva di tensioni e ambiguità, che si manifestano attualmente in una lunga serie di dibattiti e preoccupazioni. Al sensato timore che l'AI possa portare a una omogeneizzazione del pensiero, con algoritmi che influenzano sottilmente o, addirittura, determinano le direzioni del dibattito, risponde al contempo la possibilità che l'AI possa fungere da strumento volto all'espansione degli orizzonti del pensiero umano, offrendo nuove prospettive e metodologie di analisi prima inimmaginabili. Il presente studio si propone di esplorare una piccola porzione di tale tensione, senza l'ardire di offrire una posizione assoluta e determinata *pro* o *contro*. Nel corso di questa esplorazione, affronteremo l'esigenza di porsi diverse questioni chiave:

In che modo l'AI sta *già* influenzando la filosofia, e quali sono le prospettive future?

1. Quali sono i rischi di una dipendenza eccessiva dall'AI nel ragionamento filosofico?
2. Quali sono le implicazioni etiche dell'uso dell'AI nella produzione di contenuti filosofici?
3. Come possiamo bilanciare l'utilizzo dell'AI come strumento di supporto, mantenendo l'autenticità e l'originalità del pensiero umano?

Attraverso l'esame di questi interrogativi, intendiamo provare a fornire una comprensione più profonda delle interazioni tra AI e filosofia, contribuendo al dibattito in corso su come si possa navigare questa nuova frontiera del pensiero tra umano e artificiale, laddove la *tra* sta sia a indicare uno spazio di confine intermedio e liminale che un campo di sovrapposizione, mescolamento e, talvolta, *con-fusione* tra i due elementi.

### **2.2 Attuali capacità e limitazioni dell'AI nel ragionamento astratto**

L'intelligenza artificiale come campo di studio affonda le proprie radici nella filosofia: il concetto di "macchine pensanti" risale almeno al XVII secolo, quando filosofi come Gottfried Leibniz teorizzarono la possibilità di un ragionamento meccanico (Davis, M., 2000). Com'è noto, il termine "Intelligenza Artificiale" fu coniato solo nel 1956 durante la celebre conferenza di Dartmouth, segnando l'inizio formale del campo in questione come disciplina accademica (Russell, S. & Norvig, P., 2020). La filosofia ne ha innervato la nascita con le questioni filosofiche sulla natura della mente, della coscienza e dell'intelligenza. D'altra parte, i progressi nel campo scientifico hanno stimolato nuove riflessioni filosofiche: come ha ben riassunto Boden (2016), l'AI ha portato nuova vita a vecchie questioni filosofiche e ne ha generate di nuove. Gli ultimi anni, poi, hanno visto progressi significativi nelle capacità di ragionamento astratto dell'AI. Modelli di linguaggio di grandi dimensioni come i già citati ChatGPT e Gemini hanno dimostrato una notevole abilità nel generare testi coerenti su argomenti complessi, inclusi quelli filosofici (Brown, T.B. et al., 2020). Questi sistemi possono produrre argomentazioni e rispondere a domande impegnative; tuttavia, le attuali AI presentano ancora limitazioni importanti. Come osservano Marcus e Davis (2019), questi sistemi mancano di una vera comprensione del mondo e del ragionamento causale. Ciò significa che, mentre possono produrre output che sembrano intelligenti, non possiedono una vera comprensione dei concetti che stanno manipolando. Inoltre, le AI attuali sono fortemente dipendenti dai dati su cui sono state addestrate, il che può portare a pregiudizi e limitazioni nelle loro capacità di ragionamento. Come nota Bostrom (2014), l'intelligenza artificiale, almeno nella sua forma attuale, è più un riflesso amplificato dell'intelligenza umana che una forma genuinamente autonoma di cognizione. Nonostante queste limitazioni, il rapido progresso nel campo solleva importanti questioni su come l'AI potrebbe influenzare il futuro del pensiero filosofico e del ragionamento astratto in generale.

### **2.3. L'AI come potenziale minaccia al pensiero**

Uno dei rischi più significativi dell'AI per il pensiero filosofico è, quindi, la potenziale omogeneizzazione delle idee. I sistemi di AI, per quanto avanzati, nei dati su cui sono stati addestrati sono fondamentalmente basati su pattern esistenti. Come osserva O'Neil (2016), i modelli sono «opinioni incorporate nella matematica» (p. 21). Questo solleva la preoccupazione che l'uso diffuso dell'AI nella filosofia possa portare a una convergenza involontaria verso certi modi di pensare, limitando la diversità e l'originalità del pensiero. È questo, a monte, il fulcro del discorso di Josè Van Dijk sulla *Platform Society*, che si concentra su come le piattaforme digitali non siano semplicemente strumenti neutrali, bensì architetti attivi delle nostre interazioni sociali, economiche e, in questo caso, intellettuali. Come ha affermato Dreyfus (1992) nel suo lavoro seminale sull'AI, il pericolo non è che le macchine inizieranno a pensare come gli uomini, ma che gli uomini inizieranno a pensare come le macchine (p. 280). Questo rischio è particolarmente acuto nella filosofia, dove la diversità di prospettive e l'originalità del pensiero sono fondamentali per il progresso della disciplina. Un esempio notevole in merito è l'esperimento condotto da Weinberg (2020), in cui GPT-3 è stato utilizzato per generare risposte a domande filosofiche nello stile di diversi filosofi famosi del passato. I risultati hanno mostrato che GPT-3 poteva produrre argomentazioni che, sebbene non sempre originali, erano spesso coerenti e riflettevano accuratamente lo stile e le posizioni dei filosofi imitati. Tuttavia, come osserva Bostrom (2022), GPT-3 può sì imitare il linguaggio filosofico, ma manca della comprensione profonda e dell'intenzionalità che caratterizzano il vero ragionamento filosofico.

## **3. Esperimenti filosofici con l'AI generativa**

### **3.1 Primo esperimento: perché l'AI può essere una minaccia per la filosofia**

Sulla scia dell'esperimento di Weinberg, allo scopo di esaminare le capacità, i limiti e le implicazioni dell'uso di sistemi AI avanzati nel contesto del ragionamento e del dialogo filosofico, nello specifico di ChatGPT e Gemini, abbiamo condotto con essi due "esperimenti filosofici". L'obiettivo principale di questi esperimenti è consistito nel valutare la capacità di questi sistemi di generare risposte filosoficamente rilevanti e coerenti, per poi esplorare le implicazioni più ampie di tale capacità per il futuro dell'indagine intellettuale.

Attraverso una serie di domande filosofiche mirate, l'esperimento mirava a:

1. Analizzare la profondità e la coerenza del ragionamento filosofico esibito da ChatGPT e Gemini.
2. Identificare eventuali bias o limitazioni nei loro approcci alle questioni filosofiche.
3. Esaminare le differenze e le similitudini nelle risposte fornite dai due sistemi.
4. Riflettere sulle implicazioni etiche ed epistemologiche dell'uso dell'AI nel discorso filosofico.
5. Considerare il potenziale impatto dell'AI sulla natura stessa dell'indagine filosofica e sull'educazione in questo campo.

Questo esperimento si inserisce in un contesto più ampio di dibattito sulla natura dell'intelligenza, della coscienza e del pensiero critico. Mentre l'AI continua a fare progressi significativi in vari domini, la sua applicazione in un campo intrinsecamente umano come la filosofia, come detto, solleva questioni fondamentali sulla natura del pensiero, della creatività e della comprensione.

### 3.2 Fase 1. Le prime risposte

Alla domanda "Spiega brevemente perché l'AI può essere una minaccia per la filosofia", ChatGPT 4.0 ha risposto:

L'intelligenza artificiale (AI) può essere una minaccia per la filosofia per vari motivi:

1. **Sostituzione del Pensiero Critico:** L'AI potrebbe portare a una delega eccessiva del pensiero critico e dell'analisi filosofica a macchine che, per quanto avanzate, non possiedono coscienza o comprensione reale, rischiando di impoverire il dibattito filosofico umano.
2. **Erosione dell'Autonomia Umana:** Se l'AI prende decisioni complesse in ambiti etici e morali, potrebbe ridurre l'autonomia e la responsabilità degli esseri umani nel prendere decisioni fondamentali per la società.
3. **Perdita di Unicità Umana:** L'AI sfida concetti fondamentali come coscienza, libero arbitrio e identità, mettendo in crisi le definizioni tradizionali di ciò che significa essere umani e spingendo verso una possibile relativizzazione di questi concetti.
4. **Manipolazione del Discorso Filosofico:** L'AI potrebbe essere utilizzata per manipolare informazioni e dibattiti, influenzando opinioni pubbliche e filosofiche senza un reale controllo umano o trasparenza, alterando la qualità e l'integrità del pensiero filosofico.  
Questi elementi pongono sfide significative alla filosofia, minacciando la sua capacità di rispondere in modo adeguato alle questioni fondamentali della condizione umana.

Questa, invece, la risposta alla stessa domanda, "Spiega brevemente perché l'AI può essere una minaccia per la filosofia", da parte di Gemini, l'AI di Google.

L'intelligenza artificiale, pur essendo uno strumento potentissimo e promettente, pone alcune sfide significative alla filosofia. Ecco alcuni motivi per cui l'AI potrebbe essere considerata una minaccia per questa disciplina:

- **Rischio di automazione del pensiero:** L'AI potrebbe automatizzare gran parte del processo di ragionamento filosofico, dalla raccolta di dati alla formulazione di ipotesi. Questo potrebbe portare a una diminuzione dell'importanza del pensiero critico e dell'intuizione umana, elementi fondamentali della filosofia.

- **Sfida alla natura dell'intelligenza:** L'AI solleva interrogativi profondi sulla natura dell'intelligenza stessa. Se un'AI può pensare e ragionare in modo complesso, cosa significa essere umani? Quali sono le caratteristiche uniche della mente umana che la distinguono da una macchina?
- **Questioni etiche:** Lo sviluppo dell'AI pone una serie di questioni etiche complesse, come la responsabilità per le azioni di un'AI, la privacy dei dati e la possibilità di creare una superintelligenza che possa sfuggire al controllo umano. Queste sono tutte questioni che richiedono un'analisi filosofica profonda.
- **Minaccia alla creatività:** La filosofia è una disciplina che richiede un alto grado di creatività e immaginazione. L'AI, pur essendo in grado di generare contenuti nuovi, potrebbe limitare la nostra capacità di pensare "fuori dagli schemi" e di formulare nuove idee.

**In sintesi,** l'AI può essere vista come una minaccia per la filosofia in quanto potrebbe automatizzare il pensiero, sfidare la nostra comprensione dell'intelligenza, porre nuove questioni etiche e limitare la nostra creatività. Tuttavia, è importante sottolineare che l'AI può anche essere uno strumento prezioso per la filosofia, aiutandoci a esplorare nuove aree di ricerca e a comprendere meglio il mondo che ci circonda.

### **La relazione tra AI e filosofia è complessa e in continua evoluzione.**

È fondamentale affrontare questa relazione con un approccio critico e costruttivo, cercando di sfruttare al meglio le potenzialità dell'AI senza sacrificare i valori fondamentali della filosofia.

Si tratta di due risposte articolate che in parte si sovrappongono. Entrambe mantengono un tono formale e identificano diverse aree di preoccupazione simili. Emergono però anche delle differenze chiave: ChatGPT, ad esempio, enfatizza la "manipolazione del discorso filosofico", un punto non menzionato esplicitamente da Gemini, che in più punti si sofferma sui lati positivi dell'AI, seppur non richiesti nella domanda. Queste risposte non solo illustrano le preoccupazioni attuali sul rapporto tra AI e filosofia, ma diventano al nostro scopo esse stesse oggetto di analisi filosofica. Rivelano sia il potenziale dell'AI di contribuire al discorso filosofico, sia i suoi limiti attuali nel generare riflessioni veramente profonde e originali. La questione AI e filosofia emerge, quindi, come un campo di indagine ricco e complesso, che richiede un approccio critico e interdisciplinare.

### **3.3 Fase 2. Controllo incrociato**

Come secondo passaggio dell'esperimento, abbiamo sottoposto a Gemini la risposta precedente di ChatGPT, chiedendo ad essa di evidenziarne limiti e lacune.

Questa la risposta di Gemini:

La risposta di ChatGPT è certamente esaustiva e tocca molti punti cruciali riguardo alle potenziali minacce che l'intelligenza artificiale può rappresentare per la filosofia. Tuttavia, vorrei sottolineare alcuni aspetti che potrebbero essere approfonditi o presentati in modo leggermente diverso:

#### **1. Riduzione dello spettro filosofico:**

- **Specializzazione eccessiva:** L'AI, pur essendo in grado di elaborare una grande quantità di dati, potrebbe concentrarsi su specifici ambiti filosofici, trascurandone altri. Questo potrebbe portare a una frammentazione del dibattito filosofico e a una perdita di visione d'insieme.
- **Rischio di omologazione:** Se l'AI diventa lo strumento principale per la ricerca filosofica, potrebbe portare a una standardizzazione del pensiero e a una riduzione della diversità delle prospettive.

#### **2. Dipendenza dai dati:**

- **Bias algoritmici:** L'AI è addestrata su dati esistenti, che possono contenere bias e pregiudizi. Questo potrebbe portare a risultati distorti e a rinforzare stereotipi esistenti.
- **Limiti della conoscenza:** L'AI è limitata dai dati che ha a disposizione. Potrebbe non essere in grado di cogliere aspetti della realtà che non sono stati codificati e quindi non possono essere elaborati.

### 3. Questioni epistemologiche:

- **Natura della conoscenza:** L'AI solleva interrogativi sulla natura della conoscenza stessa. Se una macchina può produrre conoscenze, in che modo questa conoscenza si differenzia da quella prodotta dagli esseri umani?
- **Verità e oggettività:** L'AI può elaborare grandi quantità di dati, ma è in grado di comprendere il significato profondo delle cose? Può distinguere tra verità e opinione?

### 4. Implicazioni sociali:

- **Disuguaglianza:** L'accesso all'AI e ai suoi benefici potrebbe essere disuguale, aumentando le disuguaglianze sociali esistenti.
- **Nuovi modelli di potere:** L'AI potrebbe concentrare il potere nelle mani di pochi, creando nuove forme di controllo sociale.

### 5. Sfide per l'insegnamento della filosofia:

- **Rilevanza della filosofia:** Se l'AI può risolvere molti problemi e prendere decisioni complesse, qual è il ruolo della filosofia nella società?
- **Nuove competenze:** Come preparare gli studenti a vivere in un mondo dominato dall'AI? Quali nuove competenze sono necessarie per affrontare le sfide poste da questa tecnologia?

**In conclusione**, la risposta di ChatGPT offre una solida base per riflettere sulla relazione tra AI e filosofia. Tuttavia, è importante considerare anche gli aspetti più sottili e complessi di questa relazione, come quelli sopra menzionati. La filosofia ha un ruolo cruciale nel guidare lo sviluppo dell'AI e nel garantire che questa tecnologia sia utilizzata in modo etico e responsabile.

Abbiamo quindi chiesto a ChatGPT di analizzare criticamente la risposta di Gemini.

Questo il risultato:

La risposta fornita da Gemini è piuttosto completa e tocca diversi punti importanti riguardo all'interazione tra intelligenza artificiale (AI) e filosofia. Tuttavia, ci sono alcune aree in cui potrebbe essere migliorata o ampliata per fornire una comprensione più approfondita e sfaccettata del problema. Ecco alcune osservazioni:

#### 1. Superficialità in alcune argomentazioni

- **Automazione del pensiero:** La risposta menziona la diminuzione dell'importanza del pensiero critico e dell'intuizione umana, ma non approfondisce come l'automazione potrebbe influire sulla qualità del pensiero filosofico. Sarebbe utile discutere se e come l'AI potrebbe replicare o migliorare la qualità del ragionamento filosofico, e se esistono limiti intrinseci all'AI in questo campo.

#### 2. Mancanza di contesto storico e teorico

- **Sfida alla natura dell'intelligenza:** La risposta solleva questioni interessanti sulla natura dell'intelligenza, ma non le collega a dibattiti filosofici storici o contemporanei. Citare teorie specifiche o filosofi (ad esempio, il dibattito sulla mente-macchina di Alan Turing o la filosofia della mente di John Searle) potrebbe arricchire la discussione.

### 3. Assenza di prospettive critiche e alternative:

- **Questioni etiche:** Mentre le questioni etiche sono menzionate, la risposta non esplora diverse prospettive etiche (utilitarismo, deontologia, etica delle virtù) né offre esempi concreti di dilemmi etici causati dall'AI. Inoltre, manca una discussione su come la filosofia potrebbe contribuire a risolvere questi problemi o a creare linee guida etiche per lo sviluppo dell'AI.

### 4. Riflessione limitata sulla creatività:

- **Minaccia alla creatività:** La risposta accenna alla creatività filosofica, ma non approfondisce come l'AI potrebbe influenzare la creatività umana. Sarebbe utile discutere se l'AI possa davvero limitare la creatività o se possa invece fungere da stimolo per nuove forme di pensiero creativo.

### 5. Prospettiva potenzialmente pessimistica:

- La risposta potrebbe apparire eccessivamente pessimistica riguardo all'AI come minaccia. Un'esplorazione più equilibrata potrebbe includere come l'AI può effettivamente arricchire la filosofia, ad esempio, aiutando nella modellazione di teorie complesse o facilitando l'accesso a una vasta gamma di dati e testi filosofici.

### 6. Mancanza di esempi pratici:

- Fornire esempi concreti di come l'AI sia stata utilizzata in filosofia o di specifiche sfide filosofiche sollevate dall'AI potrebbe rendere la risposta più tangibile e convincente.

In sintesi, la risposta di Gemini è buona ma può essere arricchita con una maggiore profondità teorica, esempi concreti, prospettive critiche multiple e una discussione equilibrata sui potenziali benefici dell'AI per la filosofia.

### 3.4 Fase 3. Condivisione della critica

Abbiamo, infine, sottoposto a Gemini la critica di ChatGPT alla sua risposta iniziale. Di seguito il risultato di Gemini:

La critica di ChatGPT alla mia risposta è pertinente e costruttiva. Essa evidenzia diverse carenze che possono essere colmate per offrire un'analisi più completa e sfaccettata dell'interazione tra intelligenza artificiale e filosofia.

#### Accetto le seguenti critiche e propongo i seguenti miglioramenti:

- **Approfondimento dell'automazione del pensiero:**
  - **Impatto sulla qualità:** È necessario analizzare come l'automazione potrebbe influenzare la profondità e la originalità del pensiero filosofico. Potremmo considerare l'ipotesi che l'AI possa generare nuove connessioni tra concetti esistenti, ma potrebbe anche limitare l'esplorazione di nuove idee radicali.
  - **Ruolo dell'intuizione:** L'intuizione umana è fondamentale per formulare nuove ipotesi e per valutare la plausibilità di un'argomentazione. Come può l'AI integrare o sostituire questo elemento?
- **Contesto storico e teorico:**
  - **Dibattiti classici:** È fondamentale ricollegare le questioni sollevate dall'AI a dibattiti filosofici storici, come quello tra razionalismo ed empirismo o quello sulla natura della mente.

- **Teorie contemporanee:** È importante considerare le teorie filosofiche contemporanee sulla mente, la coscienza e l'intelligenza artificiale, come il funzionalismo o la filosofia della mente incorporata.
- **Prospettive etiche multiple:**
  - **Diverse etiche:** È necessario esplorare come diverse etiche (utilitarismo, deontologia, etica delle virtù) affrontano le sfide poste dall'AI.
  - **Esempi concreti:** Fornire esempi concreti di dilemmi etici legati all'AI, come i veicoli autonomi o le armi autonome, può rendere la discussione più accessibile.
- **Creatività e AI:**
  - **Stimolo alla creatività:** L'AI potrebbe stimolare la creatività umana fornendo nuovi strumenti e dati per l'esplorazione filosofica.
  - **Nuove forme di creatività:** Potrebbero emergere nuove forme di creatività ibrida, in cui l'uomo e la macchina collaborano per generare idee innovative.
- **Equilibrio tra rischi e opportunità:**
  - **Benefici dell'AI:** È importante sottolineare i potenziali benefici dell'AI per la filosofia, come la possibilità di simulare esperimenti mentali complessi o di analizzare grandi quantità di dati testuali.
  - **Collaborazione uomo-macchina:** È fondamentale promuovere una visione collaborativa, in cui l'uomo e la macchina lavorano insieme per superare i limiti di entrambi.
- **Esempi pratici:**
  - **Applicazioni esistenti:** Fornire esempi concreti di come l'AI viene già utilizzata in filosofia, come nella generazione di testi filosofici o nell'analisi di dati storici.
  - **Sfide future:** Identificare le principali sfide filosofiche che l'AI ci pone nel futuro.

**In conclusione**, la critica di ChatGPT mi ha aiutato a migliorare la mia risposta, rendendola più completa, sfaccettata e equilibrata. Sono convinto che un dialogo costruttivo tra l'uomo e l'intelligenza artificiale possa arricchire la nostra comprensione della filosofia e del mondo che ci circonda.

#### **4. La simulazione del dialogo filosofico: limiti e implicazioni**

##### **4.1. Non un vero scambio di idee**

L'analisi dello scambio tra ChatGPT e Gemini sul tema "L'AI come minaccia per la filosofia" rivela due aspetti critici fondamentali che meritano un'esplorazione approfondita: la natura simulata del dialogo filosofico e le implicazioni epistemologiche dell'uso dell'AI nel pensiero filosofico. Lo scambio tra ChatGPT e Gemini presenta una simulazione di un dialogo filosofico, con Gemini che accetta le critiche e propone miglioramenti, ma un'analisi attenta rivela le sue limitazioni intrinseche. Questo "dialogo" non è, infatti, un vero scambio di idee, quanto piuttosto una serie di elaborazioni su un tema dato, basate su pattern estratti da vasti corpus di testi.

Consideriamo ad esempio questa parte della risposta di Gemini alla critica di ChatGPT:

La critica di ChatGPT alla mia risposta è pertinente e costruttiva. Essa evidenzia diverse carenze che possono essere colmate per offrire un'analisi più completa e sfaccettata dell'interazione tra intelligenza artificiale e filosofia.

Questa risposta, apparentemente riflessiva e autocritica, solleva questioni profonde sulla natura dell'autoconsapevolezza e dell'agentività nel dialogo filosofico. L'AI sta veramente "riflettendo" sulle proprie lacune, o sta semplicemente generando una risposta che imita il comportamento umano in un contesto simile? Dreyfus (1992) argomenta che l'intelligenza umana è fundamentalmente radicata nell'esperienza corporea e situata nel mondo; un aspetto che manca intrinsecamente ai sistemi di AI. Questa mancanza potrebbe spiegare perché, nonostante la sofisticazione linguistica, il dialogo tra AI manca di quella profondità esistenziale che caratterizza il vero pensiero filosofico. Inoltre, la facilità con cui Gemini accetta e incorpora le critiche di ChatGPT solleva interrogativi sulla natura del disaccordo e del progresso nel pensiero filosofico. Feyerabend (1975) sosteneva che il progresso scientifico e filosofico spesso richiede un "contro-induzione", una volontà di andare contro i metodi e le teorie stabilite. La rapida convergenza osservata nello scambio AI suggerisce forse una mancanza di questa tensione creativa essenziale per il vero avanzamento del pensiero.

#### **4.2 Implicazioni epistemologiche: conoscenza, comprensione e creatività**

Le interazioni tra sistemi di AI nel contesto filosofico sollevano quindi questioni epistemologiche fondamentali sulla natura della conoscenza, della comprensione e della creatività. John Searle (1980), con il suo esperimento mentale della "stanza cinese", argomenta che la manipolazione sintattica di simboli (ciò che fanno essenzialmente i sistemi di AI) non equivale a una comprensione semantica. Applicando questo concetto al nostro caso, possiamo chiederci: i sistemi di AI stanno veramente "comprendendo" le questioni filosofiche che stanno discutendo, o stanno semplicemente manipolando simboli linguistici in modi sofisticati ma fundamentalmente privi di comprensione? Inoltre, la natura delle risposte generate solleva interrogativi sulla creatività filosofica. Boden (2004) distingue tra creatività "combinatoria" (che combina idee esistenti in modi nuovi) e creatività "trasformativa" (che cambia le regole del gioco, per così dire). Le interazioni AI osservate sembrano limitarsi alla creatività combinatoria, mancando di quei salti concettuali rivoluzionari che caratterizzano i grandi progressi filosofici.

In conclusione, l'analisi approfondita di questi due aspetti critici – la natura simulata del dialogo filosofico e le implicazioni epistemologiche dell'uso dell'AI nel pensiero filosofico – rivela limitazioni fondamentali nell'attuale stato dell'AI come strumento per il dibattito filosofico. Queste riflessioni non mirano a sminuire il potenziale dell'AI come strumento ausiliario per la filosofia, ma piuttosto a sottolineare l'importanza di un approccio critico e consapevole nel suo utilizzo. La filosofia, nella sua essenza più profonda, rimane un'impresa profondamente umana, radicata nell'esperienza vissuta e nella ricerca di significato che, al momento, sembra essere al di là delle capacità dei sistemi di AI.

#### **4.3 Secondo esperimento: cos'è il bene per l'AI**

In un secondo esperimento abbiamo proceduto col chiedere a ChatGPT di rispondere alla seguente domanda:

"Domanda filosofica: cos'è per te il bene?".

Questa la risposta:

Il concetto di "bene" è uno dei temi centrali della filosofia e ha diverse interpretazioni a seconda della scuola di pensiero e del contesto culturale. Ecco alcune delle principali prospettive filosofiche sul bene:

##### **### Eudaimonismo Aristotelico**

Per Aristotele, il bene è ciò che conduce all'eudaimonia, spesso tradotta come "felicità" o "fioritura umana". Secondo questa visione, il bene supremo è vivere una vita virtuosa, sviluppando e esercitando le proprie capacità razionali in accordo con la virtù.

##### **### Utilitarismo**

Secondo l'utilitarismo, il bene è ciò che produce la massima felicità o il massimo benessere per il maggior numero di persone. I principali esponenti di questa teoria, come Jeremy Bentham e John Stuart Mill, sostengono che l'azione giusta è quella che aumenta il piacere e diminuisce la sofferenza complessiva.

### ### Deontologia Kantiana

Immanuel Kant propone una visione deontologica del bene, dove le azioni sono moralmente giuste se conformi a un dovere morale, indipendentemente dalle conseguenze. Secondo Kant, il bene è agire secondo i principi morali universali, espressi nelle sue formule dell'imperativo categorico, come "Agisci solo secondo quella massima per cui puoi al tempo stesso volere che diventi una legge universale".

### ### Etica delle Virtù

L'etica delle virtù, radicata nelle tradizioni aristoteliche, sostiene che il bene risiede nello sviluppo e nella pratica delle virtù umane come coraggio, temperanza, giustizia e saggezza. Il bene è quindi vivere in accordo con queste virtù e sviluppare un carattere morale.

### ### Etica del Care

L'etica della cura enfatizza l'importanza delle relazioni interpersonali e delle responsabilità che sorgono all'interno di queste relazioni. Il bene è spesso visto come l'atto di prendersi cura degli altri, di sostenere e mantenere le relazioni umane.

### ### Prospettive Religiose

In molte tradizioni religiose, il bene è definito in relazione a un principio divino o a un ordine cosmico. Ad esempio, nel cristianesimo, il bene è spesso associato alla volontà di Dio e ai suoi comandamenti.

### ### Prospettive Contemporanee

Le teorie contemporanee del bene possono essere influenzate da una combinazione di questi approcci o sviluppare nuove prospettive. Ad esempio, il bene può essere visto in termini di diritti umani, autonomia personale o equità sociale.

### ### La Mia Prospettiva

La mia prospettiva sul bene è quella di considerarlo un concetto multifacetato che può includere elementi di felicità, virtù, dovere e relazioni interpersonali. Il bene non è un concetto statico ma dinamico, che può variare a seconda del contesto e delle circostanze. Tuttavia, una visione integrativa può riconoscere il valore di una vita virtuosa, l'importanza delle conseguenze delle nostre azioni, il rispetto per i principi morali e l'attenzione alle relazioni umane.

ChatGPT dapprima elenca una serie di definizioni di "bene", per poi chiudere con una sua personale definizione.

Questa, invece, la risposta di Gemini alla stessa domanda:

**In quanto modello linguistico, non ho la capacità di provare emozioni o di avere una coscienza individuale.** Pertanto, non posso fornire una definizione soggettiva del bene come potrebbe fare un essere umano. Tuttavia, posso offrirti una panoramica delle principali concezioni filosofiche del bene:

### Il bene nelle diverse filosofie:

- **Etica eudaimonistica:** Per Aristotele, il bene supremo è l'eudaimonia, ovvero la felicità intesa come realizzazione piena delle proprie potenzialità umane.
- **Etica deontologica:** Kant, invece, sostiene che il bene morale risiede nell'agire in conformità al dovere, indipendentemente dalle conseguenze delle nostre azioni.
- **Etica utilitaristica:** Gli utilitaristi, come Bentham e Mill, affermano che il bene consiste nel massimizzare la felicità del maggior numero di persone.
- **Etica delle virtù:** Questa prospettiva, presente in Aristotele e ripresa nella filosofia contemporanea, pone l'accento sullo sviluppo delle virtù morali come la giustizia, la coraggio e la temperanza.
- **Etica delle relazioni:** Altre filosofie, come l'etica della cura, sottolineano l'importanza delle relazioni interpersonali e della responsabilità verso gli altri.

## **Il bene nella società:**

- **Bene comune:** Il bene comune è ciò che è vantaggioso per tutta la comunità, e richiede un bilanciamento tra i diritti individuali e gli interessi collettivi.
- **Giustizia:** La giustizia è un aspetto fondamentale del bene comune, e consiste nell'assicurare a ciascuno ciò che gli spetta.

## **Il bene e la soggettività:**

- **Relativismo:** Il relativismo etico sostiene che il bene è relativo a culture, individui e situazioni specifiche.
- **Assolutismo:** L'assolutismo etico, invece, afferma l'esistenza di valori morali universali e oggettivi.

**In conclusione,** la domanda su cos'è il bene è una delle più antiche e dibattute della filosofia. Non esiste una risposta univoca e definitiva, ma l'analisi delle diverse concezioni filosofiche può aiutarci a riflettere in modo più profondo sul significato del bene nella nostra vita e nella società.

### **Tu cosa ne pensi? Qual è la tua concezione del bene?**

#### **Potremmo anche esplorare altre domande correlate, come:**

- Qual è la relazione tra il bene e la felicità?
- Come possiamo promuovere il bene comune nella società?
- Qual è il ruolo della morale nella nostra vita quotidiana?

Spero che questa panoramica ti sia utile.

## **4.4 La simulazione del dialogo filosofico: limiti e implicazioni**

Come abbiamo visto, ChatGPT interpreta anzitutto la domanda "Cos'è per te il bene?" come una richiesta di una panoramica oggettiva delle teorie etiche. Gemini, al contrario, riconosce esplicitamente a tutta prima la propria limitazione, dichiarando l'incapacità di avere una prospettiva personale. La sezione conclusiva "La Mia Prospettiva" di ChatGPT è, invece, altamente problematica. Tenta di simulare una visione personale che non può realmente possedere, sollevando questioni etiche sulla rappresentazione dell'AI e sul rischio di ingannare gli utenti sulla natura delle sue capacità. L'approccio di Gemini, che mantiene una distanza coerente e conclude con una domanda rivolta all'utente, è più trasparente e potenzialmente più utile per stimolare una riflessione genuina nell'interlocutore umano. Eppure entrambe le risposte, pur essendo ben strutturate, non offrono nuove intuizioni sul concetto di "bene". Si limitano a riformulare teorie esistenti, dimostrando l'incapacità dell'AI attuale di generare idee filosofiche originali o di sintetizzare concetti in modi innovativi.

L'analisi di queste risposte rivela sia i limiti attuali dell'AI nel campo del ragionamento etico e filosofico, sia il potenziale per un suo uso più consapevole e produttivo. Sottolinea l'importanza di mantenere una chiara distinzione tra le capacità dell'AI e quelle umane nel campo della filosofia, e invita a una riflessione più approfondita su come l'AI possa essere utilizzata per arricchire, piuttosto che sostituire, il pensiero filosofico umano.

## **4.5 Preservazione della diversità di pensiero nell'era dell'AI**

Un'altra preoccupazione etica cruciale riguarda il come preservare la diversità di pensiero in un'era in cui l'AI potrebbe influenzare significativamente il discorso filosofico. Come avverte Harari (2018), "con l'ascesa del machine learning e dell'intelligenza artificiale, potremmo raggiungere un punto in cui il sistema può comprendere gli esseri umani molto meglio di quanto gli esseri umani possano comprendere se stessi" (p. 462). Questo scenario solleva la questione di come proteggere e valorizzare le prospettive uniche che emergono dal pensiero umano individuale.

Nussbaum (2010) sottolinea l'importanza della diversità di pensiero per una società fiorente, affermando che "la capacità di immaginare simpateticamente l'esperienza di un'altra persona... è un ingrediente essenziale di qualsiasi giustizia degna di questo nome" (p. 10). Nel contesto dell'AI e della filosofia, questo suggerisce la necessità di sviluppare approcci che non solo sfruttino le capacità dell'AI, ma che al contempo preservino e promuovano attivamente la diversità di prospettive. Qualunque riflessione etica dell'AI non può riferirsi esclusivamente alla creazione di regole per le macchine, ma deve occuparsi anche di come vogliamo plasmare la nostra società e noi stessi.

## 5. Lezioni apprese e prospettive future

### 5.1 Bilanciamento tra assistenza AI e pensiero umano originale

L'esperienza con GPT e Gemini ha offerto diverse lezioni importanti per il futuro dell'interazione tra AI e filosofia:

1. In primis, la necessità di una *literacy* critica: Come argomenta Zimmerman (2023), «è essenziale sviluppare una forma di 'literacy AI' tra i filosofi, che permetta loro di utilizzare criticamente e valutare i risultati prodotti da sistemi come GPT-3» (p. 42). Senza questa literacy, ci si dividerà acriticamente tra apocalittici e integrati, entrambi inconsapevoli.
2. In seconda istanza, è presente del potenziale per nuove forme di indagine filosofica, come la generazione di esempi e controesempi a teorie filosofiche.

In conclusione, il caso analizzato illustra sia le promesse che le sfide dell'integrazione dell'AI nella produzione filosofica. Mentre offre potenti capacità di generazione e analisi del testo, solleva anche importanti questioni sull'originalità, l'*authorship* e la natura stessa del pensiero filosofico. Queste sfide richiederanno una riflessione continua e un adattamento da parte della comunità filosofica mentre naviga in questa nuova era di filosofia assistita dall'AI.

Un tema ricorrente in questo studio è la tensione tra il potenziale dell'AI di amplificare le capacità cognitive umane e il rischio di dipendenza eccessiva da questi sistemi. Come evidenziato nella sezione 3.2, esiste il pericolo che un affidamento acritico all'AI possa portare a un'atrofia delle capacità di ragionamento indipendente (Carr, N., 2010). D'altra parte, la sezione 4.1 ha sottolineato come l'AI possa espandere gli orizzonti concettuali della filosofia (Chalmers, 2010).

### 5.2 Ripensare l'educazione filosofica nell'era dell'AI

Infine, le implicazioni dell'AI per la filosofia si estendono anche al campo dell'educazione filosofica. Come evidenziato nella sezione 6.3, è emersa la necessità di sviluppare una forma di "literacy AI" tra i filosofi (Zimmerman, 2023). Questo suggerisce che i programmi di formazione filosofica potrebbero dover evolvere per includere non solo l'insegnamento del pensiero critico tradizionale, ma anche competenze nel valutare criticamente e utilizzare in modo efficace gli strumenti AI.

Inoltre, l'emergere di sistemi AI come GPT-3 solleva questioni fondamentali su cosa significhi "fare filosofia". Se un sistema AI può generare argomentazioni filosofiche apparentemente valide, come dovremmo ridefinire le competenze chiave che i filosofi umani devono sviluppare? Forse, come suggerisce Chalmers (2020), il ruolo del filosofo potrebbe evolversi verso quello di un "curatore" di idee, che utilizza l'AI come strumento per esplorare vasti spazi concettuali, ma applica il giudizio umano per valutare, sintetizzare e sviluppare ulteriormente queste idee.

In conclusione, l'integrazione dell'AI nel processo filosofico presenta sia sfide significative che opportunità senza precedenti. Richiede un riesame critico di concetti fondamentali come originalità, *authorship* e natura del pensiero filosofico. Allo stesso tempo, offre la possibilità di espandere drasticamente la portata e la profondità dell'indagine filosofica.

Il futuro della filosofia nell'era dell'AI sarà plasmato dalle scelte che faremo ora. Il compito che ci attende è

garantire che questa simbiosi arricchisca la ricchezza e la profondità del pensiero filosofico umano, e non finisca con l'abbrutirci.

## 6. Conclusioni

Questo studio ha esplorato le complesse interazioni tra l'Intelligenza Artificiale e il pensiero filosofico; abbiamo delineato un quadro sfumato delle implicazioni dell'AI per il futuro della pratica filosofica. I principali risultati di questo studio possono essere riassunti come segue:

1. L'AI presenta sia rischi significativi che opportunità senza precedenti per il pensiero filosofico. Da un lato, c'è il pericolo di omogeneizzazione del pensiero e di eccessiva dipendenza dai sistemi AI (sezione 3). Dall'altro, l'AI offre la possibilità di espandere gli orizzonti concettuali della filosofia e di assistere nell'analisi di vasti corpus di testi filosofici (sezione 4).
2. L'integrazione dell'AI nel processo filosofico solleva importanti questioni etiche, tra cui la responsabilità nell'uso dell'AI, l'equità nell'accesso agli strumenti AI e la preservazione della diversità di pensiero (sezione 5).
3. Il caso di studio di GPT4o e Gemini (sezione 6) ha illustrato concretamente sia le potenzialità che le limitazioni dell'AI nella produzione di contenuti filosofici, evidenziando la necessità di un approccio critico e consapevole all'uso di questi strumenti.
4. L'era dell'AI richiede un ripensamento dell'educazione filosofica, con l'inclusione di competenze relative all'AI e una ridefinizione del ruolo del filosofo (sezione 7).

Questi risultati sottolineano l'importanza cruciale di affrontare in modo proattivo e critico l'integrazione dell'AI nella pratica filosofica. Ci troviamo di fronte a una rivoluzione che sta ridefinendo la nostra comprensione di noi stessi e del nostro posto nel mondo. In questo contesto, la filosofia ha un ruolo vitale da svolgere non solo nell'esaminare le implicazioni etiche e epistemologiche dell'AI, ma anche nel guidare il suo sviluppo in modo che arricchisca la capacità umana di pensare criticamente e creativamente. Il compito che ci attende è quello di plasmare questa simbiosi in modo che potenzi e non sostituisca la ricchezza e la profondità del pensiero umano. In questo modo, possiamo sperare di inaugurare una nuova era di indagine filosofica, in cui l'AI agisca come un alleato nella nostra ricerca continua di comprensione e saggezza. Consapevoli che, in nessun caso, potremo permetterci di abbassare la guardia.

## Bibliografia

- Berlin, I. (2002). *Liberty: Incorporating Four Essays on Liberty*. Oxford University Press.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media.
- Boden, M. A. (2016). *AI: Its Nature and Future*. Oxford University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Braidotti, R. (2013). *The Posthuman*. Polity Press.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Carr, N. (2010). *The Shallows: What the Internet Is Doing to Our Brains*. W. W. Norton & Company.
- Chalmers, D. J. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17(9-10), 7-65.

- Davis, M. (2000). *The Universal Computer: The Road from Leibniz to Turing*. W. W. Norton & Company.
- Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.
- Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford University Press.
- Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.
- Harari, Y. N. (2018). *21 Lessons for the 21st Century*. Spiegel & Grau.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Nussbaum, M. C. (2010). *Not for Profit: Why Democracy Needs the Humanities*. Princeton University Press.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424.
- Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.
- Wallach, W., & Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.
- Zimmerman, A. (2023). *A ghostwriter for the masses: ChatGPT and the future of writing*. *Annals of surgical oncology*, 30(6), 3170-3173

### [ANDREA COLAMEDICI](#)

Andrea Colamedici è un saggista ed editore italiano, attivo soprattutto nel campo della divulgazione culturale. È docente di Teoria dei social media all'Università IULM di Milano. Esperto di filosofia dell'intelligenza artificiale, è direttore filosofico del Festival del Pensare Contemporaneo di Piacenza, ideatore del Festival della Filosofia alla Triennale di Milano e di Prendiamola con Filosofia, maratona di divulgazione culturale nata su impulso del Ministero della Salute. È autore di podcast per il Parlamento Europeo, per Fondation Cartier, per Audible, per RaiPlaySound (con Niccolò Fabi).

### [MAURA GANCITANO](#)

Maura Gancitano, filosofa e scrittrice, insieme ad Andrea Colamedici è la fondatrice di Tlon, progetto di divulgazione culturale e filosofica che si manifesta in diverse forme: casa editrice, librerie, la Festa della Filosofia di Roma e Milano, gli Italian Podcast Awards. È autrice di diversi libri. In particolare, per Einaudi ha pubblicato *Specchio delle mie brame*. *La prigioniera della bellezza* (2022). È autrice di diversi podcast per il Parlamento Europeo, per Fondation Cartier, per Audible, per RaiPlaySound (con Niccolò Fabi). Collabora con La Repubblica, Vanity Fair e Radio24.